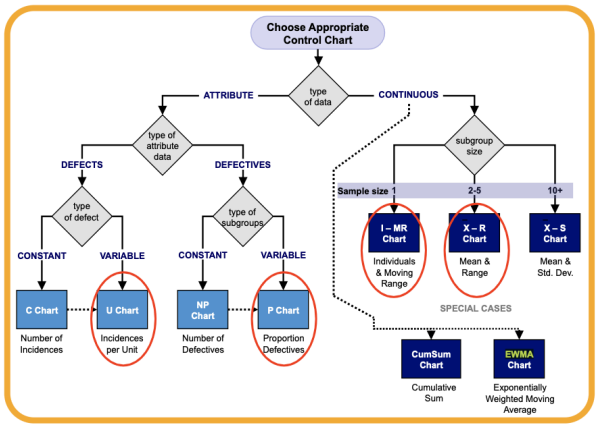
Correlation/regression explores association, not causality.; doe true cause and effect y = a + bx, b is the coefficient/slope, A is intercept/constant The R-squared value estimates, not pearson	Constant variance: Residuals versus fitted values chart Independence: Residuals versus order of data chart Normality: fat pencil plot

Exponentially Weighted Moving Average: **Small shift needs to be detected often because of autocorrelation of the output results. Used only for individuals or averages of Outputs. Infrequently used because of calculation complexity.** **Cumulative Sum**: **Same reasons as EWMA (Exponentially Weighted Moving Range) except the past data is as important as present data.**	2 most common standard transformations: square root, log **Box cox** disadvantage: Difficulty in reversing the transformation, transform power: lambda

Nonlinear relationships: Scatter Plot indicating curvature. Unusual observations in Linear Regression model. Trends of the residuals versus the Fitted Values Plot in Simple Linear Regression. Subject matter expert knowledge or team experience. If only one measurement can be taken at a time, the chart to use is the moving range chart	I-MR Charts when each measurement represents one batch.subgroup size equal to one R chart is used instead of s because it's easier to calculate S chart is More sensitive to variations. C chart count # of defects per item using a constant sample size When the lower limit of a c-chart is less than zero: change to 0 A U Chart can do everything a C Chart can, U Chart does not need a constant sample size like the sample size like the C Chart.



Control Charts indicate Special Causes being either assignable causes or patterns.

The following rules are applicable for both variable and Attribute Data to detect Special Causes.

These four rules are the only applicable tests for Range (R), Moving Range (MR) or Standard Deviation (S) charts.

- **One point more than 3 Standard Deviations from the Center Line.**
- **6 points in a row all either increasing or all decreasing.**
- **14 points in a row alternating up and down.**
- **9 points in a row on the same side of the center line.**

These remaining four rules are only for variable data to detect Special Causes.

- **2 out of 3 points greater than 2 Standard Deviations from the Center Line on the same side.**
- **4 out of 5 points greater than 1 Standard Deviation from the Center Line on the same side.**
- **15 points in a row all within one Standard Deviation of either side of the Center Line.**
- **8 points in a row all greater than one Standard Deviation of either side of the Center Line.**

Multi vari studies: within-piece variation (positional), piece-to-piece variation (cyclical), and time-to-time variation (temporal) Classic causes: Positional, cyclical, temporal	A tool that graphically displays patterns of variation. Used to identify possible X's or families of variation. Helps in screening factors by using graphical techniques. Used to assess capability, stability and graphical relationships between X's and Y's.

o (Special Causes) More than 3 standard deviations from centerline o Shift in the process mean o Indicating a trend o Indicating a non-random pattern o Indicating a worsening variation o Indicating a shift in the mean or degradation of variation o Indicating a dramatic improvement of variation o Indicating a severe worsening of variation	Understand the Normal distribution, nonnormal distributions (skewed, kurtosis, multi-modal, granularity). kurtosis (Leptokurtic, Platykurtic). 1) the random error, 2) the standard error, and, 3) the Type I error and Type II error Error in sampling, bias, error in measurement, measurement validity

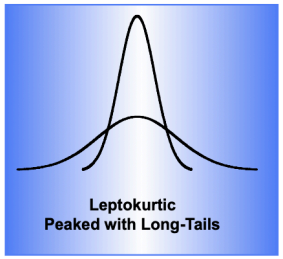
Problem Solving
- Eliminate defective products or services.
- Reduce cycle time of handling transactional processes.

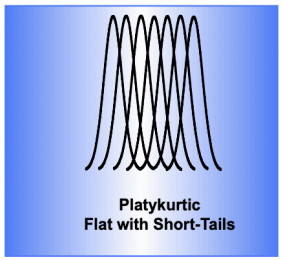
Optimizing
- Mathematical model is desired to move the process response.
- Opportunity to meet differing customer requirements (specifications or VOC).

Robust Design
- Provide consistent process or product performance.
- Desensitize the output response(s) to input variable changes including NOISE variables.
- Design processes knowing which input variables are difficult to maintain.

Screening
- Past process data is limited or statistical conclusions prevented good narrowing of critical factors in Analyze Phase.

When it rains it PORS!





Design of Experiments **(DOE) is a scientific method of planning and conducting an experiment that will yield the true cause and effect relationship between the X variables and the Y variables of interest.**

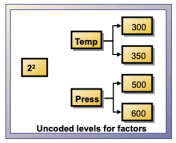
DOE allows the experimenter to study the effect of many input variables that may influence the product or process *simultaneously*, as well as possible interaction effects (for example synergistic effects).

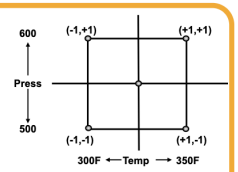
The end result of many experiments is to describe the results as a mathematical function.

$$Y = f(x)$$

The goal of DOE is to find a design that will produce the information required at a minimum cost.

Properly designed DOE's are more efficient experiments.



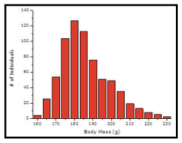


T	P	T*P
-1	-1	+1
+1	-1	-1
-1	+1	-1
+1	+1	+1

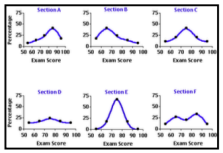
Coded levels for factors

Four experimental runs:
- **Temp = 300, Press = 500**
- **Temp = 350, Press = 500**
- **Temp = 300, Press = 600**
- **Temp = 350, Press = 600**

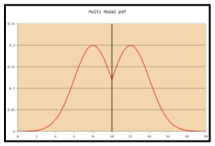
1 Skewed



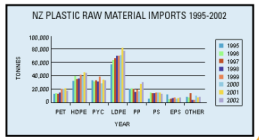
2 Kurtosis



3 Multi-Modal

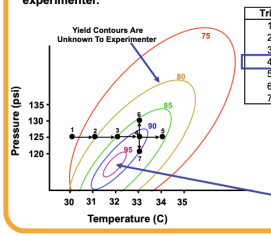


4 Granularity



One Factor at a Time **(OFAT) is an experimental style but not a planned experiment or DOE.**

The graphic shows yield contours for a process that are unknown to the experimenter.

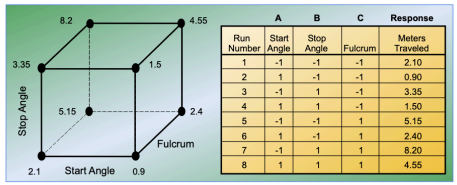


Trial	Temp	Press	Yield
1	125	30	74
2	125	31	80
3	125	32	85
4	125	33	92
5	125	34	86
6	130	33	85
7	120	33	90

Optimum identified with OFAT

True Optimum available with DOE

Consider a 2^3 design on a catapult...

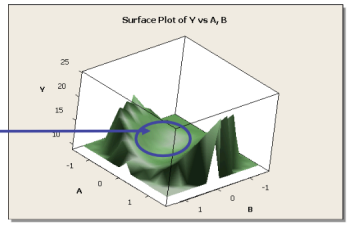


Run Number	A: Start Angle	B: Stop Angle	C: Fulcrum	Response: Meters Traveled
1	-1	-1	-1	2.10
2	1	-1	-1	0.90
3	-1	1	-1	3.35
4	1	1	-1	1.50
5	-1	-1	1	5.15
6	1	-1	1	2.40
7	-1	1	1	8.20
8	1	1	1	4.55

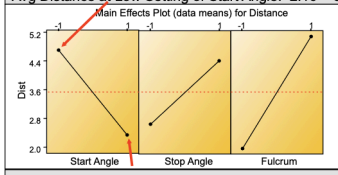
What are the inputs being manipulated in this design?
How many runs are there in this experiment?

What are the differences between DOE modeling and physical models?
- **A physical model is known by theory using concepts of physics, chemistry, biology, etc...**
- **Physical models explain outside area of immediate project needs and include more variables than typical DOE models.**
- **DOE describes only a small region of the experimental space.**

The objective is to minimize the response. The **physical model** is not important for our business objective. The **DOE Model** will focus in the region of interest.



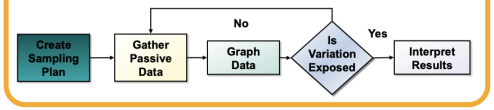
Avg Distance at Low Setting of Start Angle: 2.10 + 3.35 + 5.15 + 8.20 = 18.8/4 = 4.70



Avg. distance at High Setting of Start Angle: 0.90 + 1.50 + 2.40 + 4.55 = 9.40/4 = 2.34

Run #	Start Angle	Stop Angle	Fulcrum	Distance
1	-1	-1	-1	2.10
2	1	-1	-1	0.90
3	-1	1	-1	3.35
4	1	1	-1	1.50
5	-1	-1	1	5.15
6	1	-1	1	2.40
7	-1	1	1	8.20
8	1	1	1	4.55

The most common types of DOE's are:
- ***Fractional Factorials***
  - **4-15 input variables**
- ***Full Factorials***
  - **2-5 input variables**
- ***Response Surface Methods*** **(RSM)**
  - **2-4 input variables**

Sampling Plans should encompass all three types of variation: Within, Between and Temporal.

1. Create Sampling Plan
2. Gather Passive Date
3. Graph Data
4. Check to see if Variation is Exposed
5. Interpret Results